

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
7 February 2002 (07.02.2002)

PCT

(10) International Publication Number
WO 02/10210 A2

(51) International Patent Classification⁷: **C07K 14/415**

patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(21) International Application Number: PCT/EP01/09892

(22) International Filing Date: 28 August 2001 (28.08.2001)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant (for all designated States except US): **BAYER AKTIENGESELLSCHAFT** [DE/DE]; 51368 Leverkusen (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **TIETJEN, Klaus** [DE/DE]; Am Alten Broich 98, 40764 Langenfeld (DE). **WEIDLER, Marcus** [DE/DE]; St.-Engelbert-Strasse 31, 51519 Odenthal (DE).

(74) Common Representative: **BAYER AKTIENGESELLSCHAFT**; 51368 Leverkusen (DE).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)
- of inventorship (Rule 4.17(iv)) for US only

Published:

- upon request of the applicant, before the expiration of the time limit referred to in Article 21(2)(a)
- without international search report and to be republished upon receipt of that report
- without classification; title and abstract not checked by the International Searching Authority
- with sequence listing part of description published separately in electronic form and available upon request from the International Bureau

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/10210 A2

(54) Title: POLYPEPTIDES FOR IDENTIFYING NEW HERBICIDALLY ACTIVE COMPOUNDS

(57) Abstract: The invention relates to a method of identifying plant-specific polypeptides and nucleic acids encoding them which are suitable as sites of action for finding herbicides, to the use of the polypeptides identified for identifying new, herbicidally active compounds, and methods of finding modulators of these polypeptides. Likewise, the invention relates to the use of the polypeptides in assay methods for identifying herbicidally active compounds.

Polypeptides for identifying new herbicidally active compounds

The invention relates to a method of identifying plant-specific polypeptides and nucleic acids encoding them which are suitable as sites of action for finding
5 herbicides, to the use of the polypeptides which have been identified for identifying new herbicidally active compounds, and to methods of finding modulators of this polypeptide. Equally, the invention relates to the use of the plant polypeptides in assay methods for identifying herbicidally active compounds.

10 Herbicides have great importance in agriculture to avoid undesired plant growth by using herbicides. In modern agriculture, the use of herbicides constitutes an imperative factor for safeguarding yields and profits. This is where herbicides must meet increasingly high demands with regard to their efficacy, costs and above all their ecofriendliness. There is therefore a constant demand for new substances,
15 known as lead structures, which can be developed into even more potent and even more ecofriendly new herbicides.

To date, only a few molecular sites of action, known as targets, play a key role for the action of herbicidal compounds. Three quarters of the entire herbicide market are
20 dominated by just 5 targets, which are the sites of action of these herbicides: acetolactate synthase, elongases for very long-chain fatty acids, enolpyruvylshikimate-3-phosphate synthase, the photosystem II and the auxin signal cascade. The remaining quarter of the market comprises just 6 further important targets: acetal-coenzyme A carboxylase, glutamine synthase, photosystem I, phytoene
25 desaturase, protoporphyrinogen oxidase and tubulin. Herbicides for all of these targets have been known for over 20 years. During this period, herbicides with other, new targets have not gained market relevance. This situation leads to a thorough knowledge and exploitation of these targets in the search for new herbicidally active lead structures. At the same time, however, the use of new targets is extremely
30 important for an innovation in the search for new lead structures for the development of novel and superior herbicides.

- 2 -

To date it is generally customary to search for new lead structures in greenhouse tests. However, such tests require a good deal of labour and are expensive. The number of the substances, which can be tested in the greenhouse, is accordingly limited. However, even after suitable automation for increasing the throughput, greenhouse screening does not allow any findings as to whether substances may be directed against a new target. This must be determined in very complex subsequent experiments.

10 An alternative to the search for lead structures which is nowadays generally customary is what is known as high-throughput screening or ultra-high-throughput screening (HTS or UHTS). This method, which was first established in pharmaceutical research, makes possible the automation of in-vitro assays for the search for lead structures for given targets. At the same time, it has been made possible to provide a high number of test substances by methods such as, for example, combinatorial chemistry. Thus, a multiplicity of methods has been developed as to how specific targets can be assayed by (U)HTS. The target-based search for lead structures for agricultural applications with the aid of (U)HTS does not differ from that for pharmaceutical applications and is therefore firmly established at present.

(U)HTS makes it possible to test the action of several hundreds of thousands of substances on a specific target within a few days. However, existing experience in industry shows that it is not possible to find a lead structure for each new target, at least not at present. It is therefore necessary to test a multiplicity of targets in order to identify suitable targets in addition to new herbicidal substances.

All of the five abovementioned herbicide targets which dominate the market, and most of the remaining targets, are only found in plants but not in animals. This is no coincidence but is due to the advantageous properties of such active compounds. Thus, there is only little danger of a toxic effect on humans and the environment in

- 3 -

plant-specific targets. This can be proved by comparing the two targets acetolactate synthase and protoporphyrinogen oxidase. At the beginning of the 80s, highly effective and innovative compounds were discovered for both targets, initially without knowing the target. A series of herbicides were quick to reach the market in the case of the plant-specific target acetolactate synthase, so that acetolactate synthase is currently ranked third among the herbicide targets. Even though a very large variety of herbicides which act on protoporphyrinogen oxidase, which is also found in animals, is now known, the unfavourable toxicology of these products has as yet not led to an important commercial product.

10

Toxicological studies are complicated and expensive. As a rule, these studies are only performed when a certain basic development of new lead structures has already taken place. Even so, the research expenses up to this point are quite considerable. It is therefore advantageous to minimize the toxic effect of new herbicides, which is due to the target, right at the beginning. This can be achieved by simply using those targets for the search for lead structures which are found only in plants, but not in animals.

15

Especially advantageous targets for new herbicides are searched for in essential biosynthetic pathways. Thus, for example, the biosynthesis of isoprenoids, building blocks of carotenoids and of plastoquinone and chlorophyll, are imperative for the growth of plants. The inhibition of a step in this plant-specific biosynthetic pathway, also known as the 1-desoxyxylulose-5-phosphate pathway, leads to the death of a plant (DE 199 35 967). The knowledge of the plant specificity of specific metabolic pathways is currently fundamental knowledge in plant biochemistry (see, for example, B. B. Buchanan, W. Gruissem and R. L. Jones (Editors); "Biochemistry & Molecular Biology of Plants", American Society of Plant Physiologists, Rockville, MD, USA; 2000), even when it remains partially unclear which role certain proteins take on in the plant, and whether corresponding proteins or those with an equivalent task are also found in, for example, mammals.

20

25

30

- 4 -

Each new candidate herbicide must meet a number of criteria before it can be approved, and the choice of a suitable target is the first step in this search.

5 It is helpful to consider the existing genome information which is now available to the public, and to take note of some key criteria of herbicidal active compounds:

1. An active compound should be sufficiently selective and produce a herbicide which should be specific, or at least very selective, for plants (with regard to humans or animals).
- 10 2. An active compound should attack proteins or else genes which are imperative for the growth or the viability of the undesired plants, and
- 15 3. something should be known about the function of the target protein or target gene so that an assay and high-throughput screens can be established.

It is furthermore important for choosing suitable targets that the probability of identifying a new lead structure is considerably higher when the target has a natural binding property for ligands of low molecular weight. This is in contrast to, for example, individual protein components of large complexes with many subunits. The interference of protein-protein interactions by small ligands is less possible and requires, in principle, larger active compounds whose production costs are then frequently higher, so that a meaningful use of these active compounds as herbicides is made substantially more difficult. Targets with small natural ligands are, for example, enzymes, receptors and channels. Moreover, enzymes, receptors and channels can frequently be assayed more easily in assay methods (HTS or UHTS) than other proteins.

30 A possibility of recognizing plant-specific new targets is to test the enzymes or receptors and channels involved in plant-specific metabolic pathways or signal chains one after the other, using present-day biochemical knowledge (B. B. Buchanan, W.

Gruissem and R. L. Jones (Editors); "Biochemistry & Molecular Biology of Plants", American Society of Plant Physiologists, Rockville, MD, USA; 2000). However, this route carries the risk of overlooking important properties of the proteins.

5 While new routes for, for example, based on sequence information in the field of antibiotic research have already been described (see, for example, Molly B. Schmid, Novel approaches to the discovery of antimicrobial agents, Curr. Opin. Chemical Biol., 2, 529-534, 1998.), a method of identifying suitable targets for the search for herbicides on the basis of existing data from sequencing work is as yet not available.

10 It was therefore the object of the present invention to describe a method which is suitable for identifying, in an efficient and reliable fashion, those nucleic acids or polypeptides encoded by them from among sequence information available in public databases, which can be used for the search for new herbicidal active compounds as
15 plant-specific sites of action which can be obtained by a screening method. The object of the present invention was also to identify and to describe suitable target proteins by means of the method described and to make these available for use in screening methods for the search for new active compounds.

20 The complete knowledge of the genome of *Arabidopsis*, of humans and of many other organisms now allows to filter out, by means of computer-aided comparison of the proteins encoded in the genome, those proteins which occur in one organism but not another. Thus, it is also possible to recognize plant-specific proteins whose function was hitherto unelucidated.

25 In the present context, the term "plant-specific" is understood as meaning that no similarity with proteins from animals, in particular higher animals (Metazoa; in particular Chordata) is found.

30 A series of these plant-specific proteins, however, are also found in micro-organisms (for example bacteria, fungi).

In the present invention there is now described a possibility of identifying, from publicly available information and with the aid of computer-aided methods, those proteins and the nucleic acids encoding them which are suitable for use in methods for identifying new herbicidally active compounds.

The comparison of the proteins encoded in various genomes is possible by means of a systematic alignment comparison (for example BLAST (Altschul et al., 1990), FastA (Lipman and Pearson, 1985, Pearson 1991), Search (Smith and Waterman, 1981) Hmmer (Durbin et al., 1998)) between all proteins of one organism and those of the other organism. Preferably, one organism is selected, and the presence of the homologous sequence in other organisms is then studied.

In the present invention, all of the proteins encoded in the genome of *Arabidopsis thaliana* (hereinbelow abbreviated to "Arabidopsis") are compared with all of the other sequences which are accessible in public databases. The following databases were used as source for the Arabidopsis polypeptides in the present invention:

- a) TAIR (Huala et al., 2001), which is a searchable relational database comprising information related to *Arabidopsis thaliana*, and
- b) GenBank (Benson et al., 2000), which is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences, including protein translations.

Databases which can be used for the comparison are, for example, the following:

- a) SwissProt, which is a curated protein sequence database and provides a high level of annotations (e.g. function, domains structure, variants etc.)
- b) TrEMBL and TrEMBL-New (non-redundant protein databases), which are computer-annotated supplements of Swiss Prot and contain all the translations of EMBL nucleotide sequence entries not yet integrated in SwissProt and

- 7 -

whereby TrEMBL-New is a weekly update to TrEMBL which contains the protein-coding sequences from EMBLNEW

(see Bairoch and Apweiler, 2000).

5

All of the protein-encoding genes, and/or the polypeptides encoded by them, of the databases are compared with each other (pair-wise comparison; each polypeptide with each polypeptide) in order to find homologous similarities. The rigorous Smith-Waterman algorithm is used for this purpose.

10

To assess whether a given alignment constitutes evidence for homology, it helps to know how strong an alignment can be expected from chance alone. A local alignment without gaps consists simply of a pair of equal length segments, one from each of the two sequences being compared. A modification of the Smith-Waterman or Sellers algorithms will find all segment pairs whose "scores" can not be improved by extension or trimming. These are called high-scoring segment pairs (HSPs). To analyze how high a score is likely to arise by chance, a model of random sequences is needed. For proteins, the simplest model chooses the amino acid residues in a sequence independently, with specific background probabilities for the various residues. In the limit of sufficiently large sequence lengths m and n , the statistics of HSP scores are characterized by two parameters, K and λ . Most simply, the expected number of HSPs with score at least S is given by the formula

15

20

$$E = Kmne^{-\lambda S}$$

25

which is the so called E-value for the score S . The parameters K and λ can be thought of simply as natural scales for the search space size and the scoring system respectively.

30

The measure for the similarity which is obtained is therefore an E-value (expect-value). As shown above, the E-value indicates the probability of which the existing

- 8 -

agreement between two proteins or else genes or nucleic acids is due to pure random chance. The smaller the E-value, the more significant a hit in the search. If, for example, the E-values are in the range of $1e-70$, this means that owing to the size of the database, only 10^{-70} hits would have been expected with the search sequence.

5 This also means that the results are highly significant. In the case of two identical sequences, the E-value thus progresses towards zero. In the case of two entirely unrelated sequences, the E-value converges to values greater than one.

10 In the present method according to the invention, the criterion chosen for plant specificity and thus the suitability of the polypeptide according to the present invention, the E-value was chosen such that the exponent of the E-value of a paralogous or orthologous plant amino acid sequence must exceed that of a corresponding paralogous or orthologous animal or human sequence, in as far as such an animal or human sequence exists, at least by a factor of 3. The E-value of 10^{-30} is
15 particularly suitable as limit for defining plant specificity. If the abovementioned factor decreases, it can be assumed with high probability that the homology between the plant sequence and the animal or human sequence is too high to classify a plant polypeptide as plant-specific and suitable for the use according to the invention in methods of finding herbicides.

20

The term "identity" as used in the present context refers to the number of sequence positions which are identical in an alignment. In most cases, it is indicated as a percentage of the alignment length.

25 The term "similarity" as used in the present context, in contrast, requires the definition of a similarity matrix, that is to say a measure for the degree of similarity one wishes to assume between, for example, a valine and a threonine or a leucine.

30 The term "homology" as used in the present context, in turn, refers to evolutionary relationship. Two homologous proteins have developed from a joint precursor sequence. The term does not necessarily imply identity or similarity, apart from the

fact that homologous sequences are usually more similar (or have more identical positions in an alignment) than non-homologous sequences.

5 The term "orthologues" or "orthologous" as used in the present contexts refers to a functional counterpart, for example a protein in another organism, both having developed from a shared precursor. Normally, orthologues retain a shared function. In contrast, "paralogues" are genes or proteins resulting therefrom which have originated by duplication within a genome and which have assumed different functions during evolution which may still have similarity with each other.

10

Proteins are termed orthologous when

1. they have the highest level of pair-wise similarity (compared with the identities of the two proteins with all the other proteins in other genomes) and
- 15 2. the similarity is significant ($E < 0.01$).

15

The proteins encoded in the *Arabidopsis* genome and the results of the comparison with all the other public sequences were stored in a relational database (Oracle) in the present invention.

20

Such a relational database model was presented in 1970 by Codd et al. All of the data to be processed are shown in Tables (relations) with a fixed number of columns and any desired number of lines (tuples). Data redundancies are avoided by distributing the information to individual tables. To date, this model remains the basis of most of the commercial database systems.

25

In general, the assigning of a description which is firstly correct and can secondly be searched for readily, what is known as an annotation, to each sequence constitutes a major problem in practice. An "annotation" of a sequence is the assigning of biologically relevant properties to this sequence or parts thereof.

30

- 10 -

By comparison of, possibly competing, alternative annotations in public databases and by individual corrections, a standardized annotation for each database entry has now been generated in the present invention. For example, the annotation takes such a form that the description of enzymes, receptors and channels (transporters) starts with the respective functional name, that is, for example, with "acetolactate synthase".

An annotation was assigned to the sequence in a multi-step process: first, the information content of words or terms within a sequence description were analysed and these words/terms were correspond categorized. Thus, the description "acetolactate synthase" leads to more information on a sequence than the descriptions "Unknown Protein" or "Hypothetical Protein" or "exon predicted by xgrail, quality marginal_shadowexon". This procedure first gives two categories of words/terms and, based on these categories, eventually two categories or sequence descriptions: those with a low information content and those with a high information content.

Only the sequence descriptions with a high information content are used for assigning an annotation to a sequence. These annotations obtained in this way are subsequently aligned in a suitable fashion with the annotations obtainable from TAIR. In the present invention, the TAIR annotation for a given sequence was adopted if such an annotation did exist.

This process was automated by developing suitable programs .

In a final step, the present annotations were rechecked and, if appropriate, corrected, to arrive at the final standardized annotation.

The database established within the present invention contains sequences from Arabidopsis and the relevant descriptions(annotations) and E values in question and thus makes possible an efficient and meaningful analysis of the sequence data, which

- 11 -

results in the reliable identification of suitable plant-specific targets for the purposes of the present invention.

5 All the enzymes, receptors and channels or transporters with the above-described plant-specific E-values were then filtered out from the annotations of the database according to the invention with the aid of a suitable algorithm with suitable search terms. The polypeptides found by this method are shown in Table 1. In addition to the annotation of the polypeptide whose sequence is available by means of the reference to the sequence listing in the present application, Table 1 also shows which
10 particular class of polypeptides it belongs to. Enzymes were arranged for example by classes such as "dehydrogenase" or "oxygenase". Receptors were searched for with the search term "receptor", but not "receptor kinase". Channels were searched for with the search term "channel" or "transporter". The table also contains what is known as the accession number of the sequence, in as far as it is known. The
15 accession number provides information on the database or the number in which, or under which, the polypeptide sequence in question can be found. Furthermore, the table contains references to known homologous sequences from other organisms and a reference to the SEQ ID NO. under which the sequence in question is filed in the sequence listing.

20

Table 1:

ENTRY ID NO	SEQ ID NO	DESCRIPTION	CLASS
5	1	INORGANIC PYROPHOSPHATASE, PUTATIVE SIMILAR TO SOLUBLE INORGANIC PYROPHOSPHATASE GB: AAD46520 GI:5669924 FROM [POPULUS TREMULA X POPULUS TREMULOIDES]	Phosphatase
12	2	FATTY ACID ELONGASE 3-KETOACYL-COA SYNTHASE 1 IDENTICAL TO GB:AAC99312 GI:4091810 FROM [ARABIDOPSIS THALIANA]	Synthase
33	3	CYCLIC NUCLEOTIDE AND CALMODULIN-REGULATED ION CHANNEL, PUTATIVE SIMILAR TO CYCLIC NUCLEOTIDE AND CALMODULIN-REGULATED ION CHANNEL GB:CAB40128 GI:4581201 FROM [ARABIDOPSIS THALIANA]	Channel
38	4	FLAVONOL 3-O-GLUCOSYLTRANSFERASE, PUTATIVE SIMILAR TO FLAVONOL 3-O-GLUCOSYLTRANSFERASE GB:Q40287 FROM [MANIHOT ESCULENTA]	Transferases
41	5	FLAVONOL 3-O-GLUCOSYLTRANSFERASE, PUTATIVE SIMILAR TO FLAVONOL 3-O-GLUCOSYLTRANSFERASE GB:Q40287 FROM [MANIHOT ESCULENTA]	Transferases
46	6	1-PHOSPHATIDYLINOSITOL-4-PHOSPHATE KINASE(ATPIP5K1) GI:3702691 FROM [ARABIDOPSIS THALIANA][HYPOTHETICAL PROTEIN CONTAINS SIMILARITY TO]	5-Kinase
50	7	DEHYDROGENASE GI:1922246 FROM [ARABIDOPSIS THALIANA] UNKNOWN PROTEIN SIMILAR TO PUTATIVE	Dehydrogenases
53	8	SERINE/THREONINE KINASE, PUTATIVE SIMILAR TO SERINE/THREONINE KINASE GI:7248457 FROM [LOPHOPYRUM ELONGATUM]	Kinase, Protein
57	9	FERRIC REDUCTASE LIKE TRANSMEMBRANE COMPONENT	Reductase
58	10	FERRIC REDUCTASE LIKE TRANSMEMBRANE COMPONENT	Reductase

21189	2723	RECEPTOR PROTEIN KINASE -LIKE PROTEIN SERINE/THREONINE-SPECIFIC RECEPTOR PROTEIN KINASE (EC 2.7.1.-), ARABIDOPSIS THALIANA, PIR:S71277	Kinase, Protein
21190	2724	CELLULOSE SYNTHASE CATALYTIC SUBUNIT -LIKE PROTEIN ATH-B, CELLULOSE SYNTHASE CATALYTIC SUBUNIT, ARABIDOPSIS THALIANA, EMBL:AF027174	Synthase
21200	2725	SUGAR TRANSPORTER - LIKE PROTEIN D-XYLOSE- PROTON SYMPORTER (D-XYLOSE TRANSPORTER), LACTOBACILLUS BREVIS, SWISSPROT:XYLT_LACBR	Transporter
21202	2726	UDP GLUCOSE:FLAVONOID 3-O- GLUCOSYLTRANSFERASE -LIKE PROTEIN UDP GLUCOSE:FLAVONOID 3-O-GLUCOSYLTRANSFERASE, VITIS VINIFERA, EMBL:AF000372	Transferases
21203	2727	UDP GLUCOSE:FLAVONOID 3-O- GLUCOSYLTRANSFERASE -LIKE PROTEIN UDP GLUCOSE:FLAVONOID 3-O-GLUCOSYLTRANSFERASE, VITIS VINIFERA, EMBL:AF000371	Transferases
21204	2728	UDP GLUCOSE:FLAVONOID 3-O- GLUCOSYLTRANSFERASE -LIKE PROTEIN UDP GLUCOSE:FLAVONOID 3-O-GLUCOSYLTRANSFERASE, VITIS VINIFERA, EMBL:AF000372	Transferases
21219	2729	POLYGALACTURONASE-LIKE PROTEIN	Glycosylase
21221	2730	GLUTATHIONE S-TRANSFERASE-LIKE PROTEIN	Transferases
21222	2731	PHYTOENE SYNTHASE (GB AAB65697.1)	Synthase
21232	2732	GLUTAMATE DECARBOXYLASE I (GAD I) (SP Q42521)	Decarboxylase
21241	2733	CELLULOSE SYNTHASE CATALYTIC SUBUNIT (IRX3)	Synthase
21252	2734	PHOSPHOGLUCOMUTASE-LIKE PROTEIN PHOSPHOGLUCOMUTASE, CHLOROPLAST - SPINACIA OLERACEA, EMBL:X75898	Mutase
21281	2735	PEROXIDASE ATP13A	Oxidase
21291	2736	5-METHYLTETRAHYDROPTEROYLTRIGLUTAMATE-- HOMOCYSTEINE S-METHYLTRANSFERASE	Transferases
21297	2737	PHOSPHORIBOSYLANTHRANILATE TRANSFERASE- LIKE PROTEIN	Transferases

21308old	3209	CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
21309old	3210	CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
23001old	3211	4-DIPHOSPHOCYTIDYL-2C-METHYL-D-ERYTHRITOL SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
23094old	3212	CYSTEINE SYNTHASE, CHLOROPLAST PRECURSOR (O- ACETYL SERINE SULFHYDRYLASE) (O-ACETYL SERINE (THIOL)-LYASE) (CSASE).	Synthase
34209old	3213	CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
34659old	3214	CYSTEINE SYNTHASE ATCYSC1 [ARABIDOPSIS THALIANA].	Synthase
37280old	3215	CYSTEINE SYNTHASE, MITOCHONDRIAL PRECURSOR (O-ACETYL SERINE SULFHYDRYLASE) (O- ACETYL SERINE (THIOL)-LYASE) (CSASE).	Synthase
37284old	3216	CYSTEINE SYNTHASE (O-ACETYL SERINE SULFHYDRYLASE) (O-ACETYL SERINE (THIOL)-LYASE) (CSASE).	Synthase
39272old	3217	CHLOROPHYLL B SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
40108old	3218	CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
40109old	3219	CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
42762old	3220	5'-PHOSPHORIBOSYL-5-AMINOIMIDAZOLE SYNTHETASE.	Synthase
42911old	3221	PUTATIVE CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
44492old	3222	SIMILAR TO NICOTIANA 5-EPI-ARISTOLOCHENE SYNTHASE (GB	Synthase
44907old	3223	CYSTEINE SYNTHASE (EC 4.2.99.8) 3A - ARABIDOPSIS THALIANA.	Synthase
44988old	3224	CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
45432old	3225	CYSTEINE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase
46254old	3226	3-DEOXY-D-ARABINO-HEPTULOSONATE PHOSPHATE SYNTHASE.	7- Synthase
7417old	3227	2C-METHYL-D-ERYTHRITOL 2,4-CYCLODIPHOSPHATE SYNTHASE [ARABIDOPSIS THALIANA].	Synthase

Many annotations in publicly accessible data bases occur repeatedly, i.e. for various nucleic acid or amino acid sequences. The reasons for this are, to a minor extent, erroneous and/or redundant sequences and descriptions. To a major extent, this reflects the fact that proteins with the same function do indeed occur repeatedly in the genome. These different proteins can differ from each other for example by the regulation of their expression or by their cellular localization.

Many proteins belong to particular protein families. The skilled worker can draw conclusions with regard to the type of function, and thus also the possibility of an assay method for the polypeptide in question or its biological activity, from the protein family it belongs to. A description of such families of polypeptides and genes from Arabidopsis is obtainable for example in EP-A-1 033 405, but can also be found in the literature with which the skilled worker is familiar. Corresponding related information regarding the individual targets in Table 1 can be found in the document cited or in the general literature.

The analysis carried out for the purpose of the present invention, however, provides not only the general descriptions and the descriptions which are less suitable for the choice of herbicide targets in EP-A-1 033 405, but also the specificity of the polypeptide for the plant kingdom and the groups enzyme, receptor or channel (transporter) and more specific classes of these groups to which the proteins belong. The method according to the invention thus makes it possible to identify the particular suitability of a protein as target for finding lead structures for new herbicides exclusively with the aid of the method according to the invention. The classes which the polypeptides according to the invention were assigned to comprise, inter alia, acetylases, aldolases, amidases, amylases, anhydrases, arginases, ATPases, carboxylases, carrier-proteins, cellulases, channels, chelataases, chitinases, cyclases, deaminases, decarboxylases, dehydratases, dehydrogenases, desaturases, enolases, epimerases, esterases, furanases, furanosidases, galactosidases, galacturonases, glucanases, glucosidases, glucosylases, glucuronases, glycosylases, GTPases,

5 helicases, hydrolases, hydroxylases, isomerases, kinases, LACCases, lactonases, ligases, lipases, lyases, mannosidases, maturases, methylases, mutases, nucleases, nucleosidases, nucleotidases, oxidases, oxygenases, pectases, pectosidases, peptidases, permeases, phosphatases, phosphorylases, polymerases, proteases, racemases, receptors, reductases, sulfurylases, synthases, synthetases, transferases, transporters, transcriptases, xylanases and xylosidases.

10 The polypeptides which are identified by means of the method according to the invention are therefore particularly suitable as targets for finding new herbicidal active compounds. They are particularly suitable because they

- 15 a) have no homologous counterpart in animal organisms or in humans, according to the method according to the invention (determination of E-values, alignment of data bases),
- b) were selected with a view that they are enzymes with small ligands or else receptors or channels which can, as a rule, be modulated, i.e. inhibited or activated, by small organic molecules or peptides and are therefore in principle open to being influenced by an active compound, and
- 20 c) owing to the assignment to particular groups, make it possible for the skilled worker to select in a direct and obvious fashion assay methods which are suitable for the particular classes of polypeptides. To this end, the skilled worker can rely on the current literature or exploit the assay methods
- 25 described in the present application.

30 Subject-matter of the present invention is therefore furthermore the use of polypeptides found with the aid of the method according to the invention or of the nucleic acids encoding these polypeptides in methods for finding modulators of the polypeptides according to the invention or for finding new herbicidal compounds.

Subject-matter of the present invention is in particular the use of one of the polypeptides of SEQ ID NO: 1 to SEQ ID NO: 3227 in methods for finding modulators of these polypeptides or for finding new herbicidal compounds.

5

The subject-matter of the present invention is furthermore the use of polypeptides which exert at least the biological activity of one of the polypeptides according to the invention and which encompass an amino acid sequence which has at least 60% identity, preferably 80% identity, especially preferably 90% identity, very especially preferably 97% identity, with a sequence of SEQ ID NO: 1 to SEQ ID NO: 3227 over its entire length in methods for finding modulators of the polypeptides or for finding new herbicidal active compounds.

10

The degree of identity of the amino acid sequences is determined for example with the aid of the program BLASTP + BEAUTY Version 2.0 4. (Altschul et al., 1997).

15

Preferred polypeptides which are used in the methods for finding modulators of the polypeptides according to the invention are those of SEQ ID NO: 1 to SEQ ID NO: 3227.

Based on the genetic code, a nucleic acid sequence encoding these polypeptides can be deduced in a simple fashion from the amino acid sequences of the polypeptides according to the invention, which amino acid sequences are shown in the sequence listing.

20

Such deduced nucleic acids can be used as probes and/or primers for detection and/or isolation of related polynucleotide sequences in different organisms, preferably in plants, through hybridization. Depending on the stringency of the conditions under which these probes and primers are used, polynucleotides exhibiting a wide range of similarity to those shown in Table 1 can be detected or isolated. "Stringency" as used herein is a function of probe length, probe composition (G/C content) and salt

25

30

concentration, organic solvent concentration and temperature of hybridization or wash conditions. Stringency is typically compared by the parameter T_m , which is the temperature of hybridization or wash conditions. Stringency is typically compared by the parameter T_m which is the temperature at which 50% of the complementary molecules in the hybridization are hybridized. High stringency conditions are e.g. those providing a condition of T_m 5°C to 10°C. Medium or moderate stringency conditions are those providing T_m 20°C to T_m 29°C. Low stringency conditions are those providing for a condition of T_m 40°C to T_m 48°C. The relationship of hybridization conditions to T_m (in °C) is expressed in the following equation:

10

$$T_m = 81.5 - 16.6 (\log_{10}[\text{Na}^+] + 0.41(\%G+C)) - (600/N),$$

where N is the length of the probe. This equation works well for probes comprising 14 to 70 nucleotides in length that are identical to the target sequence.

15

Subject-matter of the present invention is therefore also the use of the nucleic acids encoding the polypeptides according to the invention in methods for finding new herbicidal compounds, and of DNA constructs which encompass one of the deduced nucleic acid sequences and a homologous or heterologous promoter.

20

The term "homologous promoter" as used in the present context refers to a promoter which controls the expression of the gene in question in the original organism.

25

The term "heterologous promoter" as used in the present context refers to a promoter which has properties other than the promoter which controls the expression of the gene in question in the original organism.

30

The choice of heterologous promoters depends on whether pro- or eukaryotic cells or cell-free systems are used for expression. Examples of heterologous promoters are the cauliflower mosaic virus 35S promoter for plant cells, the alcohol dehydrogenase

promoter for yeast cells, the T3, T7 or SP6 promoters for prokaryotic cells or cell-free systems.

5 Subject-matter of the present invention is furthermore vectors comprising a nucleic acid encoding a polypeptide according to the invention or an abovementioned DNA construct. Vectors which can be used are all those phages, plasmids, phagemids, phasmides, cosmids, YACs, BACs, artificial chromosomes or particles which are suitable for particle bombardment, which are used in molecular biology laboratories.

10 Preferred vectors are pBIN (Bevan, 1984) and its derivatives for plant cells, pFL61 (Minet *et al.*, 1992) or, for example, the p4XXprom. vector series (Mumberg *et al.*) for yeast cells, pSPORT vectors (Life Technologies) for bacterial cells, lambdaZAP (Stratagene) for phages or Gateway vectors (Life Technologies) for various expression systems in bacterial cells or Baculovirus.

15 Subject-matter of the present invention is furthermore host cells comprising at least one nucleic acid encoding one of the polypeptides according to the invention or a DNA construct according to the invention or a vector according to the invention.

20 The term "host cell" as used in the present context refers to cells which do not naturally comprise the nucleic acids to be used in accordance with the invention.

Suitable host cells are prokaryotic cells, preferably *E. coli*, but also eukaryotic cells, such as cells of *Saccharomyces cerevisiae*, *Pichia pastoris*, insects, plants, frog
25 oocytes and mammalian cell lines.

The term "polypeptides" as used in the present context refers not only to short amino acid chains which are usually termed peptides, oligopeptides or oligomers, but also to longer amino acid chains which are usually termed proteins. It encompasses amino
30 acid chains which can be modified either by natural processes, such as post-

translational processing, or by chemical prior-art methods. Such modifications may occur at various sites and repeatedly in a polypeptide, such as, for example, on the peptide backbone, on the amino acid side chain, on the amino and/or the carboxyl terminal. For example, they encompass acetylations, acylations, ADP ribosylations, amidations, covalent linkages to flavins, haeme moieties, nucleotides or nucleotide derivatives, lipids or lipid derivatives or phosphatidylinositol, cyclisation, disulfide bridge formations, demethylations, cystine formations, formylations, gamma-carboxylations, glycosylations, hydroxylations, iodinations, methylations, myristoylations, oxidations, proteolytic processings, phosphorylations, selenoylations and tRNA-mediated amino acid additions.

The polypeptides to be used in accordance with the invention may exist in the form of "mature" proteins or as parts of larger proteins, for example as fusion proteins. They can furthermore exhibit secretion or leader sequences, pro-sequences, sequences which make possible simple purification, such as polyhistidine residues, or additional stabilizing amino acids.

The polypeptides to be used in accordance with the invention need not constitute complete plant proteins but may also only be fragments thereof, as long as they retain at least one biological activity of the complete plant proteins. Polypeptides which exert the same type of biological activity as one of the proteins of Table 1 are still considered as being within the scope of the present invention. In this context, it is not necessary for the polypeptides to be used in accordance with the invention to be deducible from Arabidopsis proteins. Polypeptides which correspond to proteins of, for example, the plants given hereinbelow or fragments of these proteins which can still exert their biological activity are also considered as being within the scope of the present invention: tobacco, maize, wheat, barley, oats, oil seed rape, rice, rye, soya bean, tomatoes, legumes, potato plants, *Lactuca sativa*, Brassicae, woody species, *Physcomitrella patens*.

- 239 -

In comparison with the corresponding regions of the naturally occurring polypeptides, the polypeptides according to the invention can have deletions or amino acid substitutions as long as they still exert at least one biological activity of the complete polypeptides. Conservative substitutions are preferred. Such

5 conservative substitutions encompass variations, one amino acid being replaced by another amino acid from among the following group:

1. Small aliphatic residues, unpolar residues or residues of little polarity: Ala, Ser, Thr, Pro and Gly;
- 10 2. Polar, negatively charged residues and their amides: Asp, Asn, Glu and Gln;
3. Polar, positively charged residues: His, Arg and Lys;
4. Large aliphatic unpolar residues: Met, Leu, Ile, Val and Cys; and
5. Aromatic residues: Phe, Tyr and Trp.

15 The following list shows preferred conservative substitutions:

Original residue	Substitution
Ala	Gly, Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp
Gly	Ala, Pro
His	Asn, Gln
Ile	Leu, Val, Met
Leu	Ile, Val, Met
Lys	Arg,
Met	Leu, Ile

- 240 -

Original residue	Substitution
Phe	Met, Leu, Tyr, Ile, Trp
Pro	Gly
Ser	Thr
Thr	Ser
Trp	Tyr, Phe
Tyr	Trp, Phe
Val	Ile, Leu

The skilled worker knows that the polypeptides of the present invention can be obtained by various routes, for example by chemical methods such as the solid-phase method. To obtain larger protein quantities, the use of recombinant methods is recommended. The expression of a cloned gene according to the invention or fragments thereof can be effected in a series of suitable host cells which are known to the skilled worker. To this end, a nucleic acid encoding one of the polypeptides according to the invention or a DNA construct according to the invention or vector is introduced into a host cell with the aid of known methods.

10

The integration into the chromosome of the host cell, of the cloned nucleic acid according to the invention which is suitable for expressing the polypeptide according to the invention, is within the scope of the present invention. This nucleic acid or fragments thereof are preferably introduced into a plasmid, and the coding regions of the nucleic acids or fragments thereof are linked functionally to a constitutive or inducible promoter.

15

The basic steps for preparing the recombinant polypeptides according to the invention are :

20

1. Obtaining a natural, synthetic or semi-synthetic nucleic acid (DNA) which encodes a polypeptide according to the invention.

- 241 -

2. Introducing this DNA into an expression vector which is suitable for expressing the polypeptide according to the invention, either alone or as a fusion protein.
 3. Transforming a suitable host cell, preferably a prokaryotic host cell, with this expression vector.
 4. Growing this transformed host cell in a manner which is suitable for expressing the polypeptide according to the invention.
 5. Harvesting the cells and isolating the polypeptide according to the invention by suitable, known methods.
- In this context, the coding regions of the polypeptide according to the invention can be expressed for example in *E. coli* using the customary methods. Suitable expression systems for *E. coli* are commercially available, for example the expression vectors of the pET series, such as pET3a, pET23a, pET28a with His-tag or pET32a with His-tag for simple purification and thioredoxin fusion for increasing the solubility of the expressed enzyme, and pGEX with glutathione synthetase fusion, and also the pSPORT vectors, with the possibility of transferring the coding region into different vectors of the Gateway system for various expression systems. The expression vectors are transformed into λ DE3-lysogenic *E. coli* strains, for example, BL21(DE3), HMS 174(DE3) or AD494(DE3). After the initial growth of the cells under standard conditions known to the skilled worker, expression is induced by means of IPTG. After induction of the cells, incubation is carried out for 3 to 24 hours at temperatures of from 18 to 37°C. The cells are disrupted by sonication in breaking buffer (10 to 200 mM sodium phosphate, 100 to 500 mM NaCl, pH 5 to 8. The protein expressed can be purified by chromatographic methods,

- 242 -

in the case of protein expressed with His-tag by chromatography on an Ni-NTA column.

Another favourable approach is the expression of a polypeptide according to the invention in commercially available yeast strains (for example, *Pichia pastoris*) or in
5 insect cell cultures (for example Sf9 cells).

Alternatively, the polypeptides according to the invention can also be expressed in plants.

10

A rapid method of isolating the polypeptides according to the invention which are synthesized by host cells using a nucleic acid encoding them starts with the expression of a fusion protein, it being possible for the fusion moiety to be affinity-purified in a simple manner. The fusion moiety can be, for example, glutathione S-transferase. The fusion protein can then be purified on a glutathione affinity column.
15 The fusion moiety can be cleaved off by partial proteolytic cleavage for example at linkers between the fusion moiety and the polypeptide according to the invention which is to be purified. The linker can be designed such that it includes target amino acids, such as arginine and lysine residues, which define sites for trypsin cleavage. In
20 order to generate such linkers, standard cloning methods using oligonucleotides may be applied.

Other purification methods which are possible are based on preparative electrophoresis, FPLC, HPLC (for example using gel filtration columns, reverse-phase columns or mildly hydrophobic columns), gel filtration, differential
25 precipitation, ion-exchange chromatography and affinity chromatography.

The terms "isolation or purification" as used in the present context mean that the polypeptides according to the invention are separated from other proteins or other
30 macromolecules of the cell or of the tissue. Preferably, a preparation comprising the

polypeptides according to the invention is at least 10-fold concentrated and especially preferably at least 100-fold concentrated with regard to the protein content over a host cell preparation.

- 5 The polypeptides according to the invention can also be affinity-purified without fusion moieties with the aid of antibodies which bind to the polypeptides.

10 The polypeptides found here with the aid of the method according to the invention and the polypeptides which are homologous to them make possible the search for new specific herbicides; thus, ways are opened up of identifying lead structures, some of which may be completely new, with the aid of these targets. Thus, new interesting herbicides can be provided starting from such compounds which inhibit the present polypeptides.

- 15 Not only the enzymes, receptors and channels stated, but other proteins with other functions, too, can be filtered out for their plant specificity. This also applies to proteins whose function is as yet unknown.

20 Just as described above for finding new targets for herbicides, fungus- or insect-specific targets can be identified. For this purpose, the genomes of relevant phytopathogenic fungi, for example, *Magnaporthe* and many others, or insects, for example *Drosophila*, *Heliothis* and many others, are compared with the genomes of plants and animals. Thus, those enzymes, receptors and channels which are fungus-specific (and which do not occur in plants or animals) or which are insect-specific
25 (and which do not occur in plants or higher animals, that is to say Chordata, in particular humans), can be identified.

30 The search for lead structures by target-based screening has played a key role for approximately 10 years in the search for pharmaceutical active compounds. In crop protection research, the same key position has emerged somewhat later. Owing to

this high relevance, a multiplicity of methods have been developed for verifying any new target. Also included are methods of expressing the genes in relevant systems with which the skilled worker in the field of various families of proteins or classes of enzymes is generally familiar.

5

Enzymes and how they are affected by active compound candidate molecules can be measured quite generally on the basis of their enzymatic activity. The enzymatic conversion of starting materials to products can be determined in a multiplicity of ways: for example by monitoring the optical characteristics of the reaction solution (for example absorption, fluorescence, luminescence). If the enzymatic reaction cannot be monitored visually directly, the reaction can frequently be monitored by coupling with one or more further reactions, either enzymatic or non-enzymatic reactions, which can be monitored visually. As an alternative, a multiplicity of variants of binding assays have been developed which are based on measuring the binding of active compound candidate molecules to a protein. Binding assays can be carried out using radiolabeled or optically labeled detection molecules. Binding assays can also be carried out without labels, for example by methods of mass spectrometry or nuclear resonance spectrometry. This is in sharp contrast to the protein functions, which can be tested by cellular assays. Here, cells are constructed in a variety of ways which respond in a specific manner to the inhibition (or activation) of an enzyme (or receptor or channel). For example, bacteria can be constructed whose intrinsic enzyme has been switched off and was then replaced by a corresponding plant enzyme. When the action of active compound candidate molecules on the wild-type bacterial strain and the transgenic strain are compared, active compounds can be identified which relate to the plant enzyme. Cellular assays can preferably be used for assaying in particular receptors, but also channels. For example, non-plant cells can be constructed which recombinantly comprise a plant receptor and which visualize the response of the receptor to active compound candidate molecules visually. Thus, a luciferase can be expressed in receptor-mediated fashion, for example, and this luciferase can then be detected with high

10

15

20

25

30

sensitivity. Channels which are ion-selective, in particular for calcium, can be detected for example by ion-selective stains.

5 The multiplicity of possibilities of opening up enzymes, receptors and channels to screening, preferably HTS or UHTS, is described in various reviews (see, for example, J. A. Landro et al., J. Pharmacol. Toxicol. Methods 44 (2201) 273 - 289). A large number of public fora exist for the specialists working in this field, such as, for example, the "Society for Biomolecular Screening" (Danbury, CT, USA) (www.sbsonline.org), which publishes its own periodical. The annual conferences of
10 the "Society for Biomolecular Screening" reflect the current state of the art. It can therefore be said that it is currently possible to convert any desired protein into an HTS assay, it being possible for the difficulty or complexity of the assay method to vary, depending on the polypeptide.

15 Many assay systems whose aim it is to assay compounds and natural extracts are designed for high throughput numbers in order to maximize the number of substances studied within a given period. Assay systems which are based on cell-free procedures require purified or semipurified protein. They are suitable for a "first" assay, whose principal aim is to detect a potential effect of a substance on the target
20 protein.

Effects such as cell toxicity are, as a rule, ignored in these *in vitro* systems. The assay systems test both inhibitory or suppressive effects of the substances and stimulatory effects. The efficacy of a substance can be tested by concentration-dependent test
25 series. Control batches without test substances can be used for assessing the effects.

In the following text, methods shall be shown by way of example which can be exploited inter alia for finding modulators of the polypeptides according to the invention, the methods according to the invention including high-throughput
30 screening (HTS) and ultra-high throughput screening (UHTS). Both host cells and

cell-free preparations comprising the nucleic acids according to the invention and/or the polypeptides according to the invention can be used for this purpose.

5 The examples given are understood as being a nonlimiting selection of methods which are possible for use for the purpose in accordance with the invention.

Activity assays

10 In order to find modulators of the polypeptides to be used according to the invention, for example a synthetic reaction mix (for example products of the *in vitro* transcription) or a cellular component, such as a crude cell extract, or any other preparation comprising the polypeptide to be used in accordance with the invention can be incubated together with one or more optionally labeled substrates or ligands of the polypeptides in the presence or absence of a candidate molecule, which may be an agonist or antagonist. The ability of the candidate molecule of increasing or
15 inhibiting the activity of the polypeptide to be used in accordance with the invention can be seen from an increased or reduced conversion of the substrate. Molecules which lead to an increased activity of the polypeptides to be used in accordance with the invention are agonists. Molecules which lead to a reduction in the activity of the polypeptides to be used in accordance with the invention are probably inhibitors or
20 antagonists. The detection of the biological activity of the polypeptides to be used in accordance with the invention can possibly be improved by what is known as a reporter system. Reporter systems as used herein comprise, but are not limited to, colorimetrically labeled substrates which are converted into a product, or a reporter gene which responds to changes in the activity or the expression of the polypeptides
25 to be used in accordance with the invention.

Binding assays

30 In order to find modulators of the polypeptides to be used according to the invention, for example a synthetic reaction mix (for example products of the *in vitro* transcription) or a cellular component, such as a crude cell extract, or any other

preparation comprising the polypeptide to be used in accordance with the invention can be incubated together with a labeled substrate or ligand of the polypeptides in the presence or absence of a candidate molecule, which may be an agonist or antagonist. The ability of the candidate molecule of increasing or inhibiting the activity of the polypeptide to be used in accordance with the invention can be seen from an increased or reduced binding of the labeled ligand. Molecules which bind well and lead to an increased activity of the polypeptides to be used in accordance with the invention are agonists. Molecules which bind well but do not trigger the biological activity of the polypeptides to be used in accordance with the invention are probably good antagonists. The detection of the biological activity of the polypeptides to be used in accordance with the invention can possibly be improved by what is known as a reporter system. Reporter systems as used herein comprise, but are not limited to, a reporter gene which responds to changes in the activity or expression of the polypeptides to be used in accordance with the invention, or other known binding assays.

Displacement assays

A further example of a method by means of which modulators of the polypeptides to be used in accordance with the invention can be found is a displacement assay in which the polypeptides to be used in accordance with the invention and a potential modulator are contacted under suitable conditions with a molecule which is known to bind to the polypeptides to be used in accordance with the invention, such as a natural substrate or ligand, or a substrate or ligand mimetic. The polypeptides to be used in accordance with the invention can be labeled themselves, for example radiolabeled or colorimetrically labeled, so that the number of the polypeptides which are bound to a ligand or which have undergone a conversion can be determined accurately. In this manner, the efficacy of an agonist or antagonist can be determined.

- 248 -

For the purposes of molecular interaction studies using a polypeptide according to the invention, or else with polypeptide variants which have been modified by *in vitro* mutagenesis or other known methods, a known analytical system may be employed, for example by Biacore AB, Uppsala, Sweden. In this system, (i) the polypeptide according to the invention or fragments thereof can be coupled to a biochip via known chemical methods (coupling via amines, thiols, aldehydes) or affinity binding (for example Streptavidin-Biotin, IMAC), or (ii) a ligand, for example a peptide or a small molecule, can be coupled to the chip. The binding, to the immobilized molecules, of a ligand in solution can be measured physically. In the case of the Biocore Instrument, the ligand is immobilized on a sensor chip with a thin gold layer. The solution of the analyte is perfused through a micro-flow cell on the chip. The binding of the analyte to the immobilized ligand increases the local concentration at the surface, the refractive index of the medium close to the gold layer gradually increasing. This affects the interaction between free electrons (plasmons) in the metal and photons which are emitted by the instrument. These physical changes are proportional to the mass and molecular number on the chip, the ligand-analyte binding is registered in real time, thus allowing the apparent association/dissociation rate to be determined (Fivash et al. 1998). Competition experiments validate the specificity of the binding. Analogous measurements also serve to determine the polypeptide domains which are important for the binding of ligands, and to identify new, as yet unknown, ligands of the polypeptides according to the invention.

Scintillation Proximity Assay (SPA)

A possibility of identifying substances which modulate the activity of specific polypeptides according to the invention, such as, for example, receptor proteins, and polypeptides which are homologous thereto, is what is known as "Scintillation Proximity Assay" (SPA), see EP 015 473. This assay system exploits the interaction of a receptor with a radiolabeled ligand (for example a small organic molecule or a second radiolabeled protein molecule). The receptor is bound to microspheres or beads provided with scintillating molecules. As the radioactivity declines, the

scintillating substance in the microsphere is excited by the subatomic particles of the radiolabel, and a detectable photon is emitted. The assay conditions are optimized in such a way that only those particles originating from the ligand lead to a signal which originate from a ligand bound to the receptor or to the polypeptide according to the invention.

In a possible embodiment, the polypeptide according to the invention is bound to the beads, either together with, or without, interacting or binding test substances. It would also be possible to use fragments of the polypeptides according to the invention. When a binding, for example radiolabeled, ligand binds to the immobilized polypeptide according to the invention, this ligand should inhibit or cancel out an existing interaction between the immobilized polypeptide according to the invention and the labeled ligand in order to bind itself in the contact area zone. Successful binding to the polypeptide according to the invention can then be detected by means of a flash of light. Analogously, an existing complex between an immobilized polypeptide and a free, labeled ligand is destroyed by the binding of a test substance, which leads to a drop in the intensity of the flash of light which is detected. In this case, the assay system corresponds to a complementary inhibition system.

Two Hybrid System

An example of an assay system based on intact cells is what is known as the Two Hybrid System, which is particularly suitable for those polypeptides which have a suitable interaction partner in the cell - a further polypeptide or peptide. A specific example is what is known as the interaction trap. This is a genetic selection of interacting proteins in yeast (see, for example, Gyuris *et al.* 1993). The assay system is designed to detect and describe the interaction of two proteins, owing to an interaction which has taken place leading to a detectable signal.

- 250 -

Such an assay system can also be adapted to the testing of large numbers of test substances in a given period.

The system is based on the construction of two vectors, the bait vector and the prey vector. A gene encoding a polypeptide according to the invention or fragments thereof is cloned into the bait vector and then expressed as fusion protein together with the LexA protein, a DNA binding protein. A second gene encoding an interaction partner of the polypeptide in question is cloned into the prey vector, where it is expressed as fusion protein together with the B42 prey protein. Both vectors are present in a *Saccharomyces cerevisiae* host which contains copies of LexA-binding DNA 5' of a lacZ or HIS3 reporter gene. If an interaction takes place between the two fusion proteins, activation of the transcription of the reporter gene results. If the presence of a test substance results in inhibition or interference with the interaction, the two fusion proteins can no longer interact and the product of the reporter gene is no longer produced.

Calcium Imaging

Calcium imaging or signalling must be considered as a further method of detecting substances which interact with polypeptides according to the invention. This method is suitable, for example, for receptors which act as Ca^{2+} channels. Here, calcium indicators are employed with the aid of which changes in the intracellular calcium level are made detectable. Within the scope of these methods, cells which express the relevant polypeptide according to the invention are employed, and these cells are loaded with calcium indicators. Upon UV excitation, an influx of calcium caused by an HC110-R agonist, or the release of intracellular calcium, leads to a change in absorption as a function of the calcium load of the indicator. In such a system, an antagonist can be recognized by the complete or partial suppression of the calcium signal induced by the agonist (for example α -LTX). Suitable calcium indicators which are possible for this purpose are Fura-2 (Sigma) or Indo-1 (Molecular Probes).

Further calcium indicators can be excited by visible light and change their fluorescence behaviour detectably as a function of their calcium load. The indicators Fluo-3 and Fluo-4 show high affinity for calcium. Fluo-4, which has the stronger fluorescence signal, is particularly suitable for measurements in test systems where the cells are employed only at low density, as is the case for HEK293 cells. Further indicators are Rhod-2, x-Rhod-1, Fluo-5N, Fluo-5F, Mag-Fluo-4, Rhod-5F, Rhod-5N, Y-Rhod-5N, Mag-Rhod-2, Mag-X-Rhod-1, Calcium Green-1 and -2, Calcium Green-5N, Oregon Green 488 BAPTA-1, Oregon Green 488 BAPTA-2 and -5N, Fura Red, Calcein and the like.

An alternative to loading cells with calcium indicators is the recombinant expression of photoproteins in the target cells. Once these photoproteins have formed a complex with calcium ions, they react in the form of a light emission. A photoprotein which has already been used often in a large number of studies and assay systems is aequorin. In this assay method, the cells which simultaneously express the target

- 252 -

protein and the aequorin are first loaded with the luminophore coelenterazin. The apoaequorin formed by the cells forms a complex with the coelenterazin and carbon dioxide. If calcium subsequently enters the cell and binds to the complex, carbon dioxide and blue light are emitted (emission maximum ~466 nm). The light emission
5 correlates with the calcium concentration which prevails intracellularly.

Subject-matter of the present invention is therefore in particular also the use of the polypeptides of the Table 1 which have been identified with the aid of the present method in methods of finding modulators of the polypeptides according to the
10 invention.

Subject-matter of the present invention is furthermore the use of nucleic acids encoding these plant proteins, DNA constructs comprising them, host cells comprising them, or antibodies which bind to these proteins in methods of finding
15 modulators of the polypeptides according to the invention.

The term "agonist" as used in the present context refers to a molecule which accelerates or increases the activity of the protein.

20 The term "antagonist" as used in the present context refers to a molecule which slows down or prevents the activity of the protein.

The term "modulator" as used in the present context constitutes the generic term for agonist and antagonist. Modulators can be small organochemical molecules, peptides
25 or antibodies which bind to the polypeptides to be used in accordance with the invention. Furthermore, modulators can be small organochemical molecules, peptides or antibodies which bind to a molecule which, in turn, binds to the polypeptides to be used in accordance with the invention, thus influencing their biological activity. Modulators can constitute natural substrates and ligands or of

structural or functional mimetics thereof. However, the term "modulator" does not extend to the natural substrates and to ATP.

The modulators are preferably small organochemical compounds.

5

The binding of the modulators to the proteins to be used in accordance with the invention can modify the cellular processes in such a way which lead to the death of the plants treated therewith.

10 Subject-matter of the present invention are therefore also modulators which have been found with the aid of one of the polypeptides described in accordance with SEQ ID NO:1 to SEQ ID NO:3227 for identifying modulators of a polypeptide.

15 Subject-matter of the invention is furthermore the use of modulators of the polypeptides in accordance with SEQ ID NO:1 to SEQ ID NO:3227 as herbicides.

Furthermore, the present invention comprises methods of finding chemical compounds which modify the expression of the polypeptides to be used in accordance with the invention. Such "expression modulators", again, can constitute
20 growth-regulatory or herbicidal active compounds. Expression modulators can be small organochemical molecules, peptides or antibodies which bind to the regulatory regions of the nucleic acids encoding the polypeptides which are to be used in accordance with the invention. Furthermore, expression modulators can be small organochemical molecules, peptides or antibodies which bind to a molecule which,
25 in turn, binds to regulatory regions of the nucleic acids encoding the polypeptides to be used in accordance with the invention, thus influencing their expression. Expression modulators can also be antisense molecules.

- 254 -

The present invention therefore also extends to the use of modulators of the polypeptides according to the invention or of expression modulators of same as plant growth regulators or herbicides.

- 5 Subject-matter of the present invention are also expression modulators of proteins which are found with the aid of any above-described method of identifying expression modulators of the proteins.

- 10 Subject-matter of the invention is also the use of expression modulators as herbicides.

References

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) "Basic local alignment search tool." J. Mol. Biol. 215:403-410.
- 5 Bairoch A. & Apweiler R. (2000) "The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000." Nucleic Acids Res. 28:45-48.
- Benson, D.A., Karsch-Mizrachi I., Lipman D.J., Ostell J., Rapp B.A., Wheeler D.L.
10 (2000) "GenBank." Nucleic Acids Res. 28(1):15-8.
- Bevan, M. (1984), Agrobacterium vectors for plant transformation, Nucl. Acids. Res. 12 (22), 8711-8721.
- 15 Codd, E. (1970) A Relational Model For Large Shared Data Banks. Communications of the ACM, 13(6):377-387.
- Durbin, R., Eddy, S., Krogh, A. & Mitchison, G. (1998) "Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids." Cambridge
20 University Press.
- Huala, E., Dickerman, A., Garcia-Hernandez, M., Weems, D., Reiser, L., LaFond, F., Hanley, D., Kiphart, D., Zhuang, J., Huang, W., Mueller, L., Bhattacharyya, D., Bhaya, D., Sobral, B., Beavis, B., Somerville, C., and Rhee, SY. (2001) "The
25 Arabidopsis Information Resource (TAIR): A comprehensive database and web-based information retrieval, analysis, and visualization system for a model plant." Nucleic Acids Res. 29(1):102-5.
- Lipman, D.J. & Pearson, W.R. (1985) "Rapid and sensitive protein similarity
30 searches." Science 227(4693):1435-41.

- 256 -

Minet, M., Dufour, M.-E. and Lacroute, F. (1992), "Complementation of *S. cerevisiae* auxotrophic mutants by *A. thaliana* cDNAs". Plant J 2, 417-422.

- 5 Mumberg et al. (1995), "Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds". Gene 156, 119-122.

Pearson, W.R. (1991) "Searching protein sequence libraries: comparison of the sensitivity and selectivity of the Smith-Waterman and FASTA algorithms."

- 10 Genomics 11(3):635-50.

Smith, T.F. & Waterman M.S. (1981) "Identification of common molecular subsequences." J. Mol. Biol. 147(1):195-7.

Patent Claims

1. Method of identifying target proteins for herbicidally active compounds,
comprising the following steps:
5
a) alignment of a nucleic acid sequence or amino acid sequence (Group 1
sequence) from plants with a nucleic acid sequence or amino acid
sequence from non-plant organisms (group 2 sequence) using suitable
search parameters,
10
b) determination of the E-value of the group 1 sequence and a similar
group 2 sequence, and
c) selecting group 1 sequences in which the exponent of the E-value
15 exceeds that of the most similar group 2 sequence at least by a factor
of 3.
2. Method according to Claim 1, characterized in that, in a further step, those
group 1 sequences are selected which are essential for the plant and, if
20 appropriate, naturally have small ligands.
3. Method according to Claim 1 or 2, characterized in that the E-value is not
lower than 10^{-30} .
- 25 4. Use of polypeptides or of nucleic acids encoding them which are found in a
method according to one of Claims 1 to 3 in a method of identifying
modulators of these polypeptides or nucleic acids.

- 258 -

5. Use of one of the polypeptides in accordance with SEQ ID NO: 1 to SEQ ID NO: 3227 and of the nucleic acids encoding them in methods of identifying modulators of these polypeptides.
- 5 6. Method of finding a chemical compound which modulates the activity of one of the polypeptides in accordance with SEQ ID NO: 1 to SEQ ID NO: 3227, comprising the following steps:
 - 10 (a) contacting a preparation or host cell comprising the polypeptide with a chemical compound or a mixture of chemical compounds under conditions which permit the interaction of a chemical compound with the polypeptide, and
 - 15 (b) identifying the chemical compound which specifically influences the activity of the polypeptide.
7. Method of finding a chemical compound which binds to one of the polypeptides in accordance with SEQ ID NO: 1 to SEQ ID NO: 3227 and/or which displaces a natural substrate or a natural ligand, comprising the following steps:
 - 20 (a) contacting a preparation or host cell comprising the polypeptide with a chemical compound or a mixture of chemical compounds under conditions which permit the interaction of a chemical compound with the polypeptide, and
 - 25 (b) identifying the chemical compound which specifically binds to the polypeptide, and/or

- 259 -

- (c) identifying the chemical compound which specifically displaces a natural substrate or a natural ligand.
- 5 8. Method of finding a chemical compound which modulates the cellular function of one of the polypeptides in accordance with SEQ ID NO: 1 to SEQ ID NO: 3227, comprising the following steps:
- 10 (a) contacting a host cell which expresses the polypeptide with a chemical compound or a mixture of chemical compounds under conditions which permit the interaction of the chemical compound with the cell and/or the polypeptide, and
- (b) identifying the chemical compound which specifically influences the cellular function of the polypeptide.
- 15 9. Method of finding a compound which modifies the expression of the polypeptide in accordance with SEQ ID NO: 1 to SEQ ID NO: 3227, comprising the following steps:
- 20 (a) contacting a host cell expressing the polypeptide with a chemical compound or a mixture of chemical compounds,
- (b) determining the polypeptide concentration, and
- 25 (c) identifying the compound which specifically influences the expression of the polypeptide.
10. Use of a modulator of one of the polypeptides in accordance with SEQ ID NO: 1 to SEQ ID NO: 3227 as herbicide.

- 260 -

11. Herbicides which are found in a method according to Claim 6 or 7.